

Fixed points avoiding Abelian k -powers

James D. Currie*

Department of Mathematics and Statistics
University of Winnipeg
515 Portage Avenue
Winnipeg, Manitoba R3B 2E9 (Canada)
`j.currie@uwinnipeg.ca`

Narad Rampersad
Department of Mathematics
University of Liège
Grande Traverse 12 (B37)
B-4000 Liège
Belgium
`nrampersad@ulg.ac.be`

July 5, 2011

Abstract

We show that the problem of whether the fixed point of a morphism avoids Abelian k -powers is decidable under rather general conditions.

Keywords: Combinatorics on words, abelian repetitions, patterns

1 Introduction

Let Σ be a finite alphabet. Consider the following decision problem:

*The author is supported by an NSERC Discovery Grant.

Given a morphism $h : \Sigma^ \rightarrow \Sigma^*$ with an infinite fixed point w and an integer $k \geq 2$, determine if w is k -power free.*

This decidability of this problem has been well studied. For instance, Berstel [3] showed that over a ternary alphabet, there is an algorithm to determine if w is squarefree. Similarly, Karhumäki [9] showed that over a binary alphabet, there is an algorithm to determine if w is cubefree. The problem was solved in general by Mignosi and Séébold [14], who showed that there exists an algorithm for this problem for all alphabet sizes and all k . (See also the work of Krieger [12] for extensions to fractional repetitions.)

In this paper we consider the analogous question for Abelian k -power freeness. In particular, we show that for morphisms h satisfying certain rather general conditions, the following problem is decidable.

Given a morphism $h : \Sigma^ \rightarrow \Sigma^*$ with an infinite fixed point w and an integer $k \geq 2$, determine if w is Abelian k -power free.*

Dekking [6] provided sufficient conditions for a morphism h to be Abelian k -power free (i.e., h maps Abelian k -power free words to Abelian k -power free words). Carpi [4] showed the existence of an algorithm to decide if a morphism satisfying certain technical conditions is Abelian squarefree. Our decision procedure for the problem stated above is based on the idea of “templates” used in [1] and [2]. The same idea also appears in the recent work of [5].

2 Preliminaries

We freely use the usual notations of combinatorics on words and formal language theory. (See for example [7, 13].) Fix positive integer m and alphabet $\Sigma = \{1, 2, \dots, m\}$. We use \mathbb{Z} to denote the set of integers, and \mathbb{Z}^n to denote the set of $1 \times n$ matrices (i.e. row vectors) with integer entries. For $u, v \in \Sigma^*$ we write $u \sim v$ if u and v are anagrams of each other, that is, if $|u|_a = |v|_a$ for all $a \in \Sigma$. We define the Parikh map $\psi : \Sigma^* \rightarrow \mathbb{Z}^m$ by

$$\psi(w) = [|w|_1, |w|_2, \dots, |w|_m], w \in \Sigma^*.$$

In other words, $\psi(w)$ is a row vector which counts the frequencies of $1, 2, \dots, m$ in w . For $w, v \in \Sigma^*$ we have $w \sim v$ exactly when $\psi(w) = \psi(v)$. Let k be a positive integer. An **Abelian k -power** is a non-empty word of the form $X_1 X_2 \cdots X_k$ where $X_i \sim X_{i+1}$, $1 \leq i \leq k-1$.

Let a morphism $\mu : \Sigma^* \rightarrow \Sigma^*$ be fixed. It will be convenient and natural for us to make some assumptions on μ :

$$\mu(1) = 1x, \text{ some } x \in \Sigma^+ \quad (1)$$

$$|\mu(a)| > 1, \text{ for all } a \in \Sigma \quad (2)$$

It follows that if $u \in \Sigma^*$, then $|u| \leq |\mu(u)|/2$.

The **frequency matrix** of μ is the $m \times m$ matrix M such that $M_{i,j} = |\mu(i)|_j$. The i^{th} row of M is thus the Parikh vector of $\mu(i)$. For $w \in \Sigma^*$ we have

$$\psi(\mu(w)) = \psi(w)M.$$

We will need some matrix theory. A standard reference is [8, Chapter 5]. An induced norm on matrices of $\mathbb{R}_{m \times m}$ is given by

$$|M| = \sup_{v \in \mathbb{R}^m} \frac{|vM|}{|v|}$$

where $|v|$ is the usual Euclidean length of vector v . We make the additional restriction on μ that M is non-singular and that $|M^{-1}| < 1$.

Let

$$L = \{w : uwv \in \mu^n(1) \text{ for some positive integer } n, \text{ some words } u, v\}.$$

Thus L is the set of factors of the image of 1 under iteration of μ . Language L is closed under μ , and each word of L is a factor of a word of $\mu(L)$. Let $N = \max_{a \in \Sigma} |\mu(a)|$.

Lemma 2.1 *If $w \in L$ and $|w| \geq N - 1$ then we can write*

$$w = A''\mu(b)C'$$

such that A'' is a (possibly empty) suffix of $\mu(a)$ and C' is a (possibly empty) prefix of $\mu(c)$ for some $a, b, c \in \Sigma$.

Proof: The alternative is that w is an interior factor of some word of $\mu(\Sigma)$, forcing $|w| \leq N - 2$. \square

3 Ancestors and k -templates

Let k be a positive integer. A k -**template** is a $(2k)$ -tuple

$$t = [a_1, a_2, \dots, a_{k+1}, d_1, d_2, \dots, d_{k-1}]$$

where the $a_i \in \{\epsilon, 1, 2, \dots, m\}$ and the $d_i \in \mathbb{Z}^m$. We say that a word w **realizes** k -template t if a non-empty factor \mathcal{I} of w has the form

$$\mathcal{I} = a_1 X_1 a_2 X_2 a_3 \dots a_k X_k a_{k+1}$$

where $\psi(X_{i+1}) - \psi(X_i) = d_i$, $i = 1, 2, \dots, k-1$. Call \mathcal{I} an **instance** of t .

Remark 3.1 The particular k -template

$$T_k = [\epsilon, \epsilon, \dots, \epsilon, \vec{0}, \vec{0}, \dots, \vec{0}]$$

will be of interest. Word \mathcal{I} is an instance of T_k if and only if \mathcal{I} has the form

$$\mathcal{I} = X_1 X_2 \dots X_k$$

where $\psi(X_{i+1}) = \psi(X_i)$, $i = 1, 2, \dots, k-1$; in other words, if and only if \mathcal{I} is an Abelian k -power.

Let

$$t_1 = [a_1, a_2, \dots, a_{k+1}, d_1, d_2, \dots, d_{k-1}]$$

and

$$t_2 = [A_1, A_2, \dots, A_{k+1}, D_1, D_2, \dots, D_{k-1}]$$

be k -templates. We say that t_2 is a **parent** of t_1 if

$$\mu(A_i) = a'_i a_i a''_i \text{ for some words } a'_i, a''_i$$

while

$$\psi(a''_{i+1} a'_{i+2}) - \psi(a''_i a'_{i+1}) + D_i M = d_i, 1 \leq i \leq k. \quad (3)$$

Lemma 3.2 (*Parent Lemma*) Suppose that $w \in \Sigma^*$ realizes t_2 . Then $\mu(w)$ realizes t_1 .

Proof: Let w contain the factor

$$\mathcal{I} = A_1 Y_1 A_2 Y_2 \cdots A_k Y_k A_{k+1}$$

where $\psi(Y_{i+1}) - \psi(Y_i) = D_i$. For each i , write $\mu(A_i) = a'_i a_i a''_i$ and let $X_i = a''_i \mu(Y_i) a'_{i+1}$. Then

$$\mu(\mathcal{I}) = a'_1 a_1 X_1 a_2 X_2 \cdots X_{k-1} a_k X_k a_{k+1} a''_{k+1}$$

and for each i ,

$$\begin{aligned} \psi(X_{i+1}) - \psi(X_i) &= \psi(a''_{i+1} \mu(Y_{i+1}) a'_{i+2}) - \psi(a''_i \mu(Y_i) a'_{i+1}) \\ &= \psi(a''_{i+1} a'_{i+2}) - \psi(a''_i a'_{i+1}) + \psi(\mu(Y_{i+1})) - \psi(\mu(Y_i)) \\ &= \psi(a''_{i+1} a'_{i+2}) - \psi(a''_i a'_{i+1}) + (\psi(Y_{i+1}) - \psi(Y_i)) M \\ &= \psi(a''_{i+1} a'_{i+2}) - \psi(a''_i a'_{i+1}) + D_i M \\ &= d_i, \text{ by (3)} \end{aligned}$$

and $\mu(w)$ contains the instance $a_1 X_1 a_2 X_2 \cdots X_k a_{k+1}$ of t_1 . \square

Lemma 3.3 *Given a k -template t_1 , we may calculate all of its parents.*

Proof: The set of candidates for the A_i in a parent, and hence for the a'_i, a''_i is finite, and may be searched exhaustively. Since M is non-singular, a choice of values for a'_i, a''_i , together with given values d_i , determines the D_i by (3). \square

Remark 3.4 Note that not all computed values for D_i may be in \mathbb{Z}^m ; some k -templates may have no parents.

Rewriting (3),

$$D_i = (d_i + \psi(a''_i a'_{i+1}) - \psi(a''_{i+1} a'_{i+2})) M^{-1}.$$

Since the a'_i, a''_i are factors of words of $\mu(\Sigma)$, there are finitely many possibilities for $c = \psi(a''_i a'_{i+1}) - \psi(a''_{i+1} a'_{i+2})$. Let C be the (finite) set of possible values for c .

Let **ancestor** be the transitive closure of the parent relation. The D_i vectors in any ancestor of k -template T_k will have the form

$$D_i = c_q M^{-q} + c_{q-1} M^{q-1} + \cdots + c_1 M^{-1} + c_0, \quad c_j \in C, j = 0, \dots, q.$$

Let $c^* = \max\{|c| : c \in C\}$ and let $r = c^*/(1 - |M^{-1}|)$. We have

$$\begin{aligned}
& |D_i| \\
&= |c_q M^{-q} + c_{q-1} M^{q-1} + \dots + c_1 M^{-1} + c_0| \\
&\leq |c_q M^{-q}| + |c_{q-1} M^{q-1}| + \dots + |c_1 M^{-1}| + |c_0| && \text{(triangle inequality)} \\
&\leq |c_q| |M^{-q}| + |c_{q-1}| |M^{q-1}| + \dots + |c_1| |M^{-1}| + |c_0| && \text{(property of the induced norm)} \\
&\leq |c_q| |M^{-1}|^q + |c_{q-1}| |M^{-1}|^{q-1} + \dots + |c_1| |M^{-1}| + |c_0| && \text{(submultiplicativity)} \\
&\leq c^* |M^{-1}|^q + c^* |M^{-1}|^{q-1} + \dots + c^* |M^{-1}| + c^* \\
&\leq c^* \sum_{i=0}^{\infty} |M^{-1}|^i && \text{(since } |M^{-1}| < 1) \\
&= \frac{c^*}{1 - |M^{-1}|} \\
&= r.
\end{aligned}$$

Thus, the D_i lie within a ball of radius r in \mathbb{R}^n . It follows that there are only finitely many D_i 's in \mathbb{Z}^n .

Lemma 3.5 *Template T_k has finitely many ancestors.*

Proof: There are finitely many choices for the $A_i \in \{\epsilon, 1, 2, \dots, m\}$ and the D_i in any ancestor. \square

Suppose that in k -template t_1 we have $\lfloor \max_i |d_i| \rfloor = \Delta$. Let \mathcal{I} be an instance of t_1 ,

$$\mathcal{I} = a_1 X_1 a_2 X_2 a_3 \dots a_k X_k a_{k+1}$$

where $\psi(X_{i+1}) - \psi(X_i) = d_i$, $i = 1, 2, \dots, k-1$.

If $i > j$, we have

$$\begin{aligned}
|X_i| - |X_j| &= \sum_{n=1}^m (|X_i|_n - |X_j|_n) \\
&= \sum_{n=1}^m \sum_{q=1}^{i-j} (|X_{j+q}|_n - |X_{j+q-1}|_n) \\
&= \sum_{n=1}^m \sum_{q=1}^{i-j} (\psi(X_{j+q})^{(n)} - \psi(X_{j+q-1})^{(n)}) \\
&\leq \sum_{n=1}^m \sum_{q=1}^{i-j} \Delta \\
&\leq mk\Delta
\end{aligned}$$

This can be argued with the opposite inequality, showing in total that

$$||X_i| - |X_j|| \leq mk\Delta.$$

If for some i we have $|X_i| \leq N-2$, then for $1 \leq j \leq k$ we have $|X_j| \leq N-2+mk\Delta$. The greatest possible length of \mathcal{I} would then be

$$|X_i| + \sum_{j=1}^{k+1} |a_j| + \sum_{j \neq i} |X_j| \leq N-2+k+1+(k-2)(N-2+mk\Delta).$$

If $|\mathcal{I}| > N+k-1+(k-2)(N-2+mk\Delta)$ then for each i we have $|X_i| > N-2$ and repeatedly using Lemma 2.1 we can write

$$\begin{aligned} \mathcal{I} &= a_1 X_1 a_2 X_2 a_3 \dots a_k X_k a_{k+1} \\ &= a_1 a_1'' \mu(Y_1) a_2' a_2'' \mu(Y_2) a_3' \dots a_k'' \mu(Y_k) a_{k+1}' a_{k+1} \end{aligned}$$

where $\mathcal{J} = A_1 Y_1 \dots A_k Y_k A_{k+1}$ is a factor of $\mu^\omega(1)$, $\mu(A_i) = a_i' a_i''$ for each i and $X_i = a_i'' \mu(Y_i) a_{i+1}'$. It follows that parent t_2 of t_1 is realized by a factor of $\mu^\omega(1)$; moreover, instance \mathcal{J} of t_2 satisfies $|\mathcal{J}| < |\mathcal{I}|/2$.

Lemma 3.6 (*Inverse Parent Lemma*) *Suppose that \mathcal{I} is a factor of $\mu^\omega(1)$ which is an instance of t_1 , and $|\mathcal{I}| > N+k-1+(k-1)[N-2+mk\Delta]$. Then for some parent t_2 of t_1 , $\mu^\omega(1)$ contains a factor \mathcal{J} which is an instance of t_2 , and such that $|\mathcal{J}| < |\mathcal{I}|$.*

4 Decidability

4.1 Main Theorem

Theorem 4.1 *Let μ be a morphism on $\{1, 2, \dots, m\}$ and M the frequency matrix of μ . Suppose that*

$$\begin{aligned} \mu(1) &= 1x, \text{ some } x \in \Sigma^+ \\ |\mu(a)| &> 1, \text{ for all } a \in \Sigma \\ |M^{-1}| &< 1 \end{aligned}$$

and M is non-singular. It is decidable whether $\mu^\omega(1)$ is Abelian k -power free.

Proof: Calculate the set T of ancestors of T_k . By Lemma 3.5 this set is finite. Word $\mu^\omega(1)$ contains an Abelian k -power iff an instance of one of these ancestors is a factor of $\mu^\omega(1)$. For each $t = [a_1, a_2, \dots, a_{k+1}, d_1, d_2, \dots, d_{k-1}] \in T$, let $D_t = \{d_1, d_2, \dots, d_{k-1}\}$. Let $D = \cup_{t \in T} D_t$, and let $\Delta = \lfloor \max_{d \in D} |d| \rfloor$. As per Lemma 3.6, the shortest instance (if any) in $\mu^\omega(1)$ of a template of T has length at most $N + k - 2 + (k - 2)(N - 2 + mk\Delta)$. We therefore generate all the factors of $\mu^\omega(1)$ of this length, and test whether any contains an instance of one of these ancestors. \square

5 Example

In the case $m = 3$, $k = 3$, Dekking showed that the fixed point of μ contains no Abelian 3-powers, where

$$\begin{aligned}\mu(1) &= 1123 \\ \mu(2) &= 133 \\ \mu(3) &= 223.\end{aligned}$$

His method of proof was elegant, but somewhat particular to his morphisms.

Here $N = 4$. We have

$$M = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

which is non-singular. A calculation with Lagrange multipliers shows that $|M| \approx 0.8589 < 1$. Applying the approach given in our Theorem, computing in the SAGE environment, we find that $T_3 = [\epsilon, \epsilon, \epsilon, \epsilon, \vec{0}, \vec{0}]$ has 1293 parents and no “grandparents”. Since T_3 is an ancestor of itself, there are 1294 ancestors in T . Examining these ancestors, we find that $\Delta = 2$. We therefore test the factors of $\mu^\omega(1)$ of length $N + k - 2 + (k - 2)(N - 2 + mk\Delta) = 25$. None of these contains an instance of a template in T . This gives an alternate, mechanical proof of Dekking’s result.

The matrices for both of Dekking’s morphisms, for Pleasants’ morphism and for Keränen’s morphism [6, 15, 10] are invertible and satisfy $|M^{-1}| < 1$ as well as our conditions (1) and (2). It follows that the results of these different authors could also be proved via the approach of the present paper.

6 Future work

Keränen [10, 11] has constructed an Abelian 2-power free quaternary word which is the fixed point of a cyclic 85-uniform morphism. His exhaustive searches have shown that this is the shortest cyclic uniform morphism which works. One would hope that much shorter, if less symmetric, morphisms exist. The result contained here suggests a new exhaustive search, considering shorter, not necessarily symmetric morphisms. The hope is that a better proof of the existence of infinite quaternary words avoiding Abelian 2-powers will result.

References

- [1] A. Aberkane, J. Currie, A cyclic binary morphism avoiding abelian fourth powers, *Theoret. Comput. Sci.* **410** (2009), 44–52.
- [2] A. Aberkane, J. Currie, N. Rampersad, The number of ternary words avoiding abelian cubes grows exponentially, *J. Integer Seq.* **7** (2004), Article 04.2.7, 13 pages.
- [3] J. Berstel, Sur les mots sans carré définis par un morphisme. In *Proc. 6th Int'l Conf. on Automata, Languages, and Programming (ICALP)*, Lecture Notes in Computer Science **71** (1979) Springer-Verlag, 16–25.
- [4] A. Carpi, On Abelian squares and substitutions, *Theoret. Comput. Sci.* **218** (1999), 61–81.
- [5] J. Cassaigne, J. Currie, L. Schaeffer, J. Shallit, Avoiding three consecutive blocks of the same size and same sum. Preprint available at <http://arxiv.org/abs/1106.5204>.
- [6] F. M. Dekking, Strongly non-repetitive sequences and progression-free sets, *J. Comb. Theory Ser. A* **27** (1979), 181–185.
- [7] J. E. Hopcroft, J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley, Reading (1979).
- [8] R. Horn, C. Johnson, *Matrix Analysis*, Cambridge, 1985.
- [9] J. Karhumäki. On cube free ω -words generated by binary morphisms. *Disc. Appl. Math.* **5** (1983), 279–297.

- [10] V. Keränen, Abelian squares are avoidable on 4 letters, *Automata, Languages and Programming*, Lecture Notes in Computer Science **623** (1992) Springer-Verlag, 41–52.
- [11] V. Keränen, New Abelian square-free L -sequences over 4 letters, *IAS 2002*, Murmansk, Russia.
- [12] D. Krieger, Critical exponents and stabilizers of infinite words. Ph.D thesis. Available at <http://uwspace.uwaterloo.ca/handle/10012/3599> .
- [13] M. Lothaire, *Combinatorics on Words*, Encyclopedia of Mathematics and its Applications **17**, Addison-Wesley, Reading (1983).
- [14] F. Mignosi and P. Séébold. If a DOL language is k -power free then it is circular. In *Proc. 20th Int'l Conf. on Automata, Languages, and Programming (ICALP)*, Lecture Notes in Computer Science **700** (1993) Springer-Verlag, 507–518.
- [15] P. A. B. Pleasants, Non-repetitive sequences, *Proc. Cambridge Philos. Soc.* **68** (1970) 267–274.